

Sampling of Random Data Streams

Gustáv CEPCIANSKY¹, Ladislav SCHWARTZ¹

¹ Department of Telecommunications and Multimedia, University of Zilina, Univerzitná 1, 01026 Zilina, Slovakia

gcepciansky@zoznam.sk, schwartz@fel.uniza.sk

Abstract. Modern telecommunication networks work on the transmission method of common data streams in which data bursts consisting of packets that further consist of particular bits are multiplexed from various traffic sources. The larger amount of data is transmitted through a transmission medium (optical fibre), the more frequently bursts occur, and the lower amount of data, the more rarely they follow. If it is required to monitor how large amount of data is being transmitted in a network branch in order to find out, to which measure of this branch is occupied, it is not necessary to take each information unit (each packet or even each particular bit). It will do if information whether a data burst occurs in the transmission or does it not occur is taken in certain time intervals – with a certain sampling frequency. The paper deals with this sampling intervals.

Keywords

Sampling, traffic, data burst, passive optical network.

1. Introduction

There are many types of data bursts on a transmission route between two network nodes from various traffic sources. The data burst length is considered as a random variable. According to Figure 1, the burst train in a common data stream can be described as a random process $X(t)$ consisting of series of random occurring rectangular pulses with a random amplitude that only gains 2 values of $x_0 = 0$ and $x_1 = 1$ and with a random time length Θ that gains values of ν_k , $k = 1, 2, \dots$

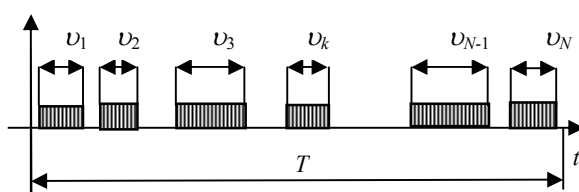


Fig. 1. Packet train in the data stream

The mean level or the average value of this random process

$$y = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^N g_k \quad (1)$$

also expresses the average traffic use of a communication channel, e.g. with which part of the total time T theoretically infinitely long the channel was occupied by the transmitted data bursts. In other words, the value y also expresses the probability that the communication channel will be in any moment of sampling t_v found as occupied.

Further, let λ be the count of seizures of the communication channel by the transmitted data bursts and μ the count of releases of the communication channel during a time unit. The ratio

$$a = \frac{\lambda}{\mu} \quad (2)$$

represents the traffic offered to the communication channel. The inverse value of μ

$$\bar{g} = \frac{1}{\mu} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N g_k \quad (3)$$

is the average burst length.

2. Theory

2.1 Random Process Characteristics

The occurrence of data bursts in the common data stream can be considered as the Markov random process because the features of such a process are fulfilled:

- a data burst occurrence is fully random,
- bursts last for a random time,
- the probability of a data burst occurrence or end of a burst during a very small time interval Δt will be proportional to the duration of this interval.

These features represent a hidden deterministic element in the random process. The consequence of this is that it is possible to explicitly derive basic characteristics – correlation and co-variation functions, dispersion and spectrum for that random process [1], [2], [3], [4].

1) Solution of Random Process

The random process gains only 2 states: S_0 – communication channel is free and S_1 – communication channel is occupied.

The process stays in the zero state S_0 if no data burst occurs during Δt , the probability of which is:

$$p_{00}(t, t + \Delta t) = 1 - \lambda \Delta t. \quad (4)$$

The process transits from the state S_0 to the state S_1 (communication channel will be seized) if a data burst occurs during Δt . The probability of this is:

$$p_{01}(t, t + \Delta t) = \lambda \Delta t. \quad (5)$$

The process stays in the state S_1 if a data burst does not finish during Δt , the probability of which is:

$$p_{11}(t, t + \Delta t) = 1 - \mu \Delta t. \quad (6)$$

The process transits from the state S_1 to the state S_0 (communication channel will be released) if a data burst finishes during Δt . The probability of that is:

$$p_{10}(t, t + \Delta t) = \mu \Delta t. \quad (7)$$

The probability that more than 1 change occurs in the state of the channel during Δt is equal to zero.

The probabilities p_{00} , p_{01} , p_{10} , p_{11} are arranged into the matrix of transition probabilities:

$$\begin{aligned} P(t, t + \Delta t) &= \begin{pmatrix} p_{00}(t, t + \Delta t) & p_{01}(t, t + \Delta t) \\ p_{10}(t, t + \Delta t) & p_{11}(t, t + \Delta t) \end{pmatrix} = \\ &= \begin{pmatrix} 1 - \lambda \Delta t & \lambda \Delta t \\ \mu \Delta t & 1 - \mu \Delta t \end{pmatrix} = \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} \Delta t = J + A \cdot t. \end{aligned} \quad (8)$$

The matrix J is the unit one. The matrix A is the matrix of intensities of transition probabilities. In general, the elements of the matrix A can be time dependent. But if it is supposed the ergodicity of the random process, e.g. the Markov process is homogenous, the time dependence does not exist.

By the solution of the basic equation of the Markov random process

$$\frac{dp(t)}{dt} = p(t) \cdot A, \quad (9)$$

it can be calculated, with which probability $p_0(t)$ the communication channel will be free and with which

probability $p_1(t)$ the communication channel will be occupied by a data burst in a time t . The equation is solved by means of Laplace transform:

$$L\{p(t)\} = F(s) = \int_0^\infty p(t) \cdot e^{-st} dt, \quad (10)$$

$$L\left\{\frac{dp(t)}{dt}\right\} = L\{p(t) \cdot A\}. \quad (11)$$

Hence, when the transform is carried out

$$F(s) = p(0) \cdot [s \cdot J - A]^{-1}. \quad (12)$$

The sought vector $p(t)$ will be obtained after the inverse Laplace transform

$$p(t) = L^{-1}\{F(s)\} = L^{-1}\{p(0) \cdot [s \cdot J - A]^{-1}\}. \quad (13)$$

First, let $s \cdot J - A$ be written in:

$$\begin{aligned} s \cdot J - A &= s \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} = \\ &= \begin{pmatrix} s + \lambda & -\lambda \\ -\mu & s + \mu \end{pmatrix} \end{aligned} \quad (14)$$

The inverse matrix:

$$\begin{aligned} [s \cdot J - A]^{-1} &= \frac{1}{s \cdot (s + \lambda + \mu)} \begin{pmatrix} s + \mu & \lambda \\ \mu & s + \lambda \end{pmatrix} = \\ &= \begin{pmatrix} \frac{s + \mu}{s \cdot (s + \lambda + \mu)} & \frac{\lambda}{s \cdot (s + \lambda + \mu)} \\ \frac{\mu}{s \cdot (s + \lambda + \mu)} & \frac{s + \lambda}{s \cdot (s + \lambda + \mu)} \end{pmatrix}. \end{aligned} \quad (15)$$

To carry out the inverse Laplace transform, the elements of the inverse matrix shall be decomposed into the sum of particle fractions. Then the basic system of equations (13) acquires the form:

$$\begin{aligned} F(s) &= p(0) \left[\frac{1}{s} \begin{pmatrix} \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \\ \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \end{pmatrix} + \right. \\ &\quad \left. + \frac{1}{s + \lambda + \mu} \begin{pmatrix} \frac{\lambda}{\lambda + \mu} & -\frac{\lambda}{\lambda + \mu} \\ -\frac{\mu}{\lambda + \mu} & \frac{\mu}{\lambda + \mu} \end{pmatrix} \right]. \end{aligned} \quad (16)$$

The result will be obtained by the inverse Laplace transform:

$$p(t) = L^{-1}\{F(s)\} = p(0) \left[\begin{pmatrix} \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \\ \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \end{pmatrix} + \right.$$

$$+ e^{-(\lambda+\mu)t} \begin{bmatrix} \frac{\lambda}{\lambda+\mu} & -\frac{\lambda}{\lambda+\mu} \\ -\frac{\mu}{\lambda+\mu} & \frac{\mu}{\lambda+\mu} \end{bmatrix}. \quad (17)$$

Hence

$$p_0(t) = \frac{1}{\lambda+\mu} [p_0(0)\mu + p_1(0)\lambda] + \frac{e^{-(\lambda+\mu)t}}{\lambda+\mu} [p_0(0)\lambda - p_1(0)\mu] \quad (18)$$

$$p_1(t) = \frac{1}{\lambda+\mu} [p_0(0)\lambda + p_1(0)\mu] + \frac{e^{-(\lambda+\mu)t}}{\lambda+\mu} [-p_0(0)\lambda + p_1(0)\mu] \quad (19)$$

These equations describe the Markov random process with 1 channel in that manner that they give the probability with which the random process will stay in one of both states in an eligible future time t .

2) Correlation and Co-variation Function

The value of the random process $X(t)$ acquires in a time t the random variable X that can acquire only 2 values: $x_0 = 0$ if a data burst is not present or $x_1 = 1$ if a data burst is present. Thus, there are 4 various combinations of products to determine the correlation function $R(\tau)$:

$$\begin{aligned} R(\tau) &= \overline{X(t) \cdot X(t+\tau)} = \\ &= \sum_{i=0}^1 \sum_{j=0}^1 x_i \cdot x_j \cdot P\{X(t) = x_i \cap X(t+\tau) = x_j\} = \\ &= x_0 \cdot x_0 \cdot P\{X(t) = x_0 \cap X(t+\tau) = x_0\} + \\ &+ x_0 \cdot x_1 \cdot P\{X(t) = x_0 \cap X(t+\tau) = x_1\} + \\ &+ x_1 \cdot x_0 \cdot P\{X(t) = x_1 \cap X(t+\tau) = x_0\} + \\ &+ x_1 \cdot x_1 \cdot P\{X(t) = x_1 \cap X(t+\tau) = x_1\} = \\ &= 1 \cdot 1 \cdot P\{X(t) = 1 \cap X(t+\tau) = 1\} \end{aligned} \quad (20)$$

If $x_0 = 0$, only the last from 4 articles has a non-zero value. As the random process is also stable and ergodic, it does not depend on the value t , so it can be $t = 0$, too. Then:

$$R(\tau) = \overline{X(t) \cdot X(t+\tau)} = \overline{X(0) \cdot X(\tau)} = \overline{P\{X(0) = 1 \cap X(\tau) = 1\}} \quad (21)$$

The probability $P\{X(0) = 1 \cap X(\tau) = 1\}$ can be get from the formula for the conditional probability:

$$\begin{aligned} P\{X(0) = 1 \cap X(\tau) = 1\} &= \\ P\{X(0) = 1\} \cdot P\{X(\tau) = 1 / X(0) = 1\} \end{aligned} \quad (22)$$

Supposing stability of the random process with probability $P\{X(0) = 1\}$, that the communication channel will be occupied at the beginning (in the time $t = 0$). The

conditional probability is equal to $P\{X(\tau) = 1 / X(0) = 1\}$, if the communication channel was occupied at the beginning, and if it will be also occupied after a time τ , it can be calculated from the equation (19), in which $p_0(0) = 0$ and $p_1 = 1$:

$$\begin{aligned} R(\tau) &= P\{X(0) = 1\} \cdot P\{X(\tau) = 1 / X(0) = 1\} \\ &= p_1(\tau) = \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} \cdot e^{-(\lambda+\mu)\tau} = \\ &= \frac{\lambda}{\mu \left(\frac{\lambda}{\mu} + 1 \right)} \left[1 + \frac{\mu}{\lambda} \cdot e^{-\left(\frac{\lambda}{\mu} + 1 \right) \mu \tau} \right] = \end{aligned} \quad (23)$$

$$\begin{aligned} &= \frac{a}{a+1} \left[1 + \frac{1}{a} \cdot e^{-(a+1) \cdot \frac{\tau}{g}} \right] \\ &= \frac{a}{a+1} \left[1 + \frac{1}{a} \cdot e^{-(a+1) \cdot \frac{\tau}{g}} \right] \end{aligned} \quad (24)$$

thereby the equations (2) and (3) were used, too.

Then the correlation function $R(\tau)$ using (22) and (24) will be:

$$R(\tau) = y \cdot \frac{a}{a+1} \left[1 + \frac{1}{a} \cdot e^{-(a+1) \cdot \frac{\tau}{g}} \right] \quad (25)$$

The following relationship holds between traffic a offered to a communication channel and their traffic use y :

$$y = (1 - p_1) \cdot a \quad (26)$$

where p_1 is the probability that a channel is occupied in the steady state:

$$p_1 = p_1(\infty) = \frac{a}{a+1} \lim_{t \rightarrow \infty} \left[1 + \frac{1}{a} \cdot e^{-(a+1) \cdot \frac{t}{g}} \right] = \frac{a}{a+1} \quad (27)$$

Substituting into (24) we obtain:

$$\begin{aligned} R(\tau) &= (1 - p_1) \cdot a \cdot \frac{a}{a+1} \left[1 + \frac{1}{a} \cdot e^{-(a+1) \cdot \frac{\tau}{g}} \right] = \\ &= \left(\frac{a}{a+1} \right)^2 \left[1 + \frac{1}{a} \cdot e^{-(a+1) \cdot \frac{\tau}{g}} \right] \end{aligned} \quad (28)$$

Co-variation function is:

$$K(\tau) = R(\tau) - m^2 \quad (29)$$

where m is the mean value (mean level) of the random process. Then

$$m^2 = y \cdot y = [(1-p_1) \cdot a]^2 = \left[\left(1 - \frac{a}{a+1}\right) \cdot a \right]^2 = \left(\frac{a}{a+1} \right)^2. \quad (30)$$

So the co-variation function will be:

$$K(\tau) = \left(\frac{a}{a+1} \right)^2 \left[1 + \frac{1}{a} e^{-\frac{(a+1)|\tau|}{g}} \right] - \left(\frac{a}{a+1} \right)^2 = \frac{a}{(a+1)^2} e^{-\frac{(a+1)|\tau|}{g}}. \quad (31)$$

If a will be expressed by means of y from (26) and (27), we obtain:

$$a = \frac{y}{1-y} \quad y < 0,$$

and then

$$K(\tau) = \frac{\frac{y}{1-y}}{\left(\frac{y}{1-y} + 1 \right)^2} e^{-\frac{\left(\frac{y}{1-y} + 1 \right) |\tau|}{g}} = y \cdot (1-y) \cdot e^{-\frac{1}{(1-y) \cdot g} |\tau|}. \quad (32)$$

Dispersion is:

$$\sigma^2 = K(0) = y(1-y). \quad (33)$$

3) Spectrum

The spectrum of a random process can be calculated by means of the Wiener-Khintchin transform which is the Fourier transform of a co-variation function:

$$S(\omega) = \int_{-\infty}^{\infty} K(\tau) \cdot e^{-j\omega\tau} d\tau. \quad (34)$$

Let in the equation (32)

$$\alpha = \frac{1}{(1-y) \cdot g}. \quad (35)$$

and σ^2 is given by (33). Then

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{\infty} \sigma^2 \cdot e^{-\alpha|\tau|} \cdot e^{-j\omega\tau} d\tau = \\ &= \sigma^2 \left[\int_0^{\infty} e^{-(\alpha-j\omega)\tau} d\tau + \int_0^{\infty} e^{-(\alpha+j\omega)\tau} d\tau \right] = \\ &= \sigma^2 \left(\frac{1}{\alpha-j\omega} + \frac{1}{\alpha+j\omega} \right) = \sigma^2 \frac{2\alpha}{\alpha^2 + \omega^2}. \end{aligned} \quad (36)$$

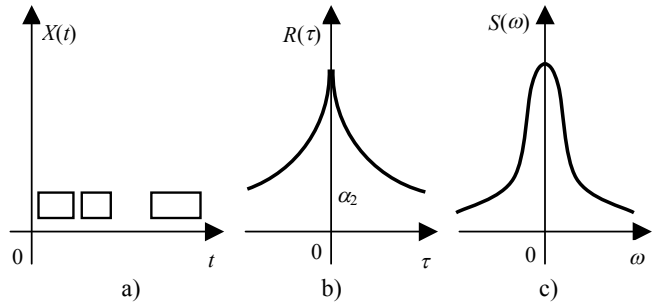


Fig. 2. Shape, co-variation function and spectrum of the random process

The data bursts train, the shape of the co-variation function and the spectrum belonging to it are drawn on Figure 2 for the comparison. The spectrum on Figure 2 c) has a shape of a probability distribution which can advantageously be used for determination of the sampling frequency.

2.2 Sampling Frequency

If the spectrum $S(\omega)$ is to be considered for a probability distribution, it must be fulfilled:

$$c. \int_{-\infty}^{\infty} S(\omega) d\omega = 1, \quad (37)$$

hence

$$c = \frac{1}{\sigma^2 \int_{-\infty}^{\infty} \frac{2\alpha}{\alpha^2 + \omega^2} d\omega} = \frac{1}{2\sigma^2 \int_{-\infty}^{\infty} \frac{dx}{1+x^2}} = \frac{1}{2\pi\sigma^2}. \quad (38)$$

So the distribution will have the shape:

$$f(\omega) = \frac{\alpha}{\pi} \cdot \frac{1}{\alpha^2 + \omega^2}, \quad (39)$$

which is the Cauchy distribution and it expresses the probability density of occurrence of amplitudes of particular spectral components. Then the elementary rectangle $f(\omega) \cdot \Delta\omega$ in Figure 3 gives the probability with which the components in the interval from ω to $\omega + \Delta\omega$ will occur in the spectrum. All spectral components in the range from $-\infty$ to $+\infty$ occur with the probability $P = 1$ because the integral from $f(\omega)$ in this range equals to 1. That means if we want to catch all changes in the course of the random process with the total certainty, it must be sampled with the infinite high frequency. Therefore, let the sampling frequency is limited by the frequency ω_s and then the probability P with which each change will be caught in the random process at this sampling frequency will be calculated:

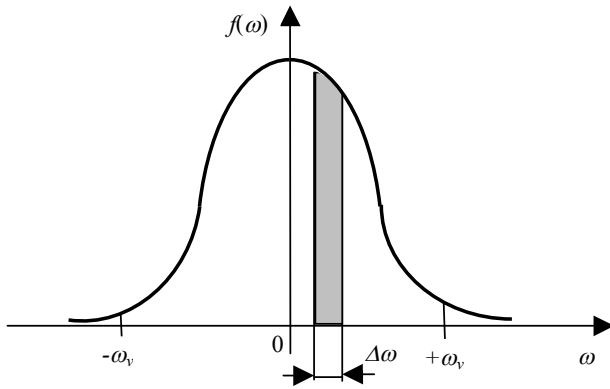


Fig. 3. Distribution of amplitudes of spectral components in the random process

$$P = \frac{\alpha}{\pi} \int_{-\omega_v}^{\omega_v} \frac{d\omega}{\alpha^2 + \omega^2} = \frac{2}{\pi} \arctg \frac{\omega_v}{\alpha} \quad (40)$$

The sampling frequency or the sampling period can be calculated from this equation for the given probability P :

$$f_v = \frac{1}{2\pi} \alpha \cdot \text{tg} \frac{\pi P}{2} = \frac{1}{2\pi(1-y)\bar{g}} \text{tg} \frac{\pi P}{2} \quad (41)$$

$$T_v = 2\pi(1-y)\bar{g} \cdot \text{cotg} \frac{\pi P}{2} \quad (42)$$

3. Experimental

3.1 Analysis of Results

The equations (41) and (42) give the sampling frequency or sampling interval, respectively in the case data bursts occur totally randomly and their length is totally random, too. It is also the case when the limit situation can occur when data bursts follow consecutively each other and the gap between them is infinitely small ($y \rightarrow 1$), or the data bursts length is infinitely short ($v \rightarrow 0$), or both cases occur simultaneously.

3.2 Practical Application

An example where this theory could be applied is the gigabit passive optical network (GPON) [5], [6], [7], [8], [9]. The GPON is an access network where more subscribers (up to 128) share the same optical fibre. The upstream communication between an optical network termination (ONT) on the subscriber side and the broadband node – optical line termination (OLT) on the network side goes through optical pulses bursts called transmission containers (T-CONTs) carrying data streams. It is necessary to hinder overlapping T-CONTs of simultaneously communicating ONTs. Therefore a time window is matched to each T-CONT from OLT. This

window defines the T-CONT length and the time instants in which an ONT may send T-CONTs with data. To determine these properly, The OLT either receives and decodes dynamic bandwidth assignment (DBA) reports coming from ONT, or monitors the traffic load on each branch in the optical network. The first method is called the status DBA and the second the non-status DBA.

The status DBA requires exchange of various control information between an OLT and an ONT in headers of data batches. The non-status DBA does not require any information exchange but it is necessary to know how often the traffic status shall be sampled on a network branch. Here the formula (41) or (42) can be helpful.

The GPON bit rate is 1 244,16 Mbit/s in the upstream. The shortest data unit is one 125 μ s long frame which is performed by 19 440 bytes. The longest data unit in the upstream is $2^{16} = 65\,536$ bytes which lasts 421,4 μ s. Let's take the mean value of these 2 figures for the average burst length:

$$\bar{g} = \frac{125 + 421,4}{2} = 273,2 \mu\text{s} \quad (43)$$

Considering $P = 0,95$, we obtain for sampling interval from (42):

$$T_v = 2\pi(1-y) \cdot 273,2 \cdot \text{cotg} \frac{0,95\pi}{2} = 65\,893 \cdot (1-y) [\mu\text{s}], \quad (44)$$

$$T_v \approx 66 \cdot (1-y) [\text{ms}]. \quad (45)$$

The longest sampling interval will be when there is no traffic load ($y = 0$):

$$T_{v,\max} = 66 \text{ ms}. \quad (46)$$

In order to avoid collision of 2 consecutive T-CONTs, the guard interval consisting of 32 bits is introduced in the upstream. They perform the guard time $32 / 1\,244\,160\,000 = 25,72 \text{ ns}$. During this time, ONT emits no energy into the optical fibre. That means there are 4-byte gaps in the 65 536-bytes bursts. That represents the traffic load of

$$y = \frac{65\,536 - 4}{65\,536} = 0,99994 \quad (47)$$

and so the shortest sampling interval will be:

$$T_{v,\min} = 66 \cdot (1 - 0,99994) = 0,004 \text{ ms} = 4 \mu\text{s} \quad (48)$$

It is necessary to take into consideration that the maximum and minimum sampling intervals $T_{v,\max}$ and $T_{v,\min}$ are calculated on condition that the data bursts occur randomly on the optical fiber and their length is also random in the range from 125 to 421,4 μ s.

Even the minimum sampling interval of 4 μ s is 5000-times longer than if each particular bit should be sampled which would give the sampling interval of $1 / 1\,244\,160\,000 = 0,8 \text{ ns}$.

4. Conclusion

Application of this theory on data streams control may save many working operations of digital microcontrollers and thus enhance their performance.

Acknowledgments

The authors gratefully acknowledge support from the VEGA project No. 1/0375/08 "The Analytic Model Home Access Network New Generation".

References

- [1] LEVIN, B., L.: Teória náhodných procesov a jej aplikácie v rádiotechnike. SNTL Praha, 1965.
- [2] SVEŠNIKOV, A. A.: Sbirka úloh z teorie pravděpodobnosti, matematické statistiky a teorie náhodných funkcí. SNTL Praha, 1971.
- [3] VENTCELOVÁ, J., S.: Teória pravdepodobnosti. Alfa-SNTL Bratislava, 1973.
- [4] PRCHAL, J.: Teorie pravděpodobnosti ve sdělovací technice. NADAS Praha, 1975.
- [5] A broadband optical access system with increased service capability using dynamic bandwidth assignment. ITU-T Recommendation G.983.4, 2001.
- [6] Gigabit-capable Passive Optical Networks (G-PON): Physical Media Dependent (PMD) layer specification. ITU-T Recommendation G.984.2, 2003.
- [7] Gigabit-capable Passive Optical Networks (G-PON): Transmission convergence layer specification. ITU-T Recommendation G.984.3, 2004.
- [8] Gigabit-capable Passive Optical Networks (G-PON): ONT management and control interface specification. ITU-T Recommendation G.984.4, 2004.
- [9] Gigabit-capable Passive Optical Networks (G-PON): Enhancement band. ITU-T Recommendation G.984.5, 2007.



Ladislav SCHWARTZ was born on 1950 in Žilina. In 1974 graduated from the University of Žilina, Department of Telecommunications with the Master's degree (MSc). From 1974 to 1991 he worked in the Research Institute of Computer Technology in Žilina as the head of the Department of Data Communications. In 1986 he was awarded the title PhD and in 1991 he became a full time teacher and research worker of the University of Žilina. In 1999 he was appointed an associate professor. His main focuses are functions, reliability, security telecommunication and computer networks and data communications.

About Authors ...



Gustav CEPCIANSKY was born in 1947. 1971 – Master of sciences (MSc) in telecommunications on University of Žilina. 1971 – 1976 – employed at Slovak Telecom as a technical development specialist. 1976 – 2002 – employed at Slovak Telecom as the head of the operational research centre. 2002 – 2009 – employed at T-Com as a technical assistance centre specialist and at the same time as a lecturer on the Telecommunication Department of the Electrical Faculty, University of Žilina. Next formation: 1984 – the 1st doctor degree (CSc). 1995 – the study stay at Technical University Supélec in Paris. 1998 – the 2nd doctor degree (PhD). 2006 – the associated professor of Electrical Faculty, University of Žilina.